

# 本周周报

解聪

2013.01.06-2013.01.12

## 本周工作

1. 经过与几位经济学老师的讨论，对问题又有了一些新的了解。我们之前一直认为是通过可视化去修改某个数据模型，比如相似性计算。但是可能领域专家更加倾向于仍然使用原先的模型。他们觉得可视化在帮助他们模型的选取上比较有帮助。比如在决定使用回归之前使用散点图先对数据做概览以确定选用哪种回归模型。

而我们的方法基本也可以做到这一点。由于数据量大，数据复杂所以一般分析师不太能确定使用何种模型。我们则可以提供一种交互式的方式使得分析师可以先一步对数据进行探索。一方面过滤数据，去除大量无关数据；另一方面通过圈选、框选等交互，指定分析师分析的上下文。然后再通过数学模型计算具体的关联规则，相关分析等。至于是否有可视分析循环探索的流程，目前还不是很清楚。

2. 类别型数据的可视分析遇到了一些理论上的问题：

- 1) 可视化的图像模式不准确？比如我们观察得到两个 Block 的图像模式互补，但是经济学家往往还需要通过统计方法来验证观察的正确性，比如使用回归模型计算参数。如果此时模型计算不能得到互补这样的模式，到底是可视化对还是模型对？

从经济学家的角度讲，我感觉大多数经济学家还是相信模型的计算结果。

- 2) 领域专家不愿意去修改数学模型。他们觉得回归这样的模型用了很久，大家都接受了。而且他们会认为我们队 Block 的框选，划选等存在误差，他们更加倾向于文本框输入参数。

- 3) 另外和顾老师聊了以后发现我的一些定义，比如：自变量，因变量，他们的相似性等还是有许多漏洞的。

2. 查看与本项目相关的文献。比如对关联规则等数据模型进行改进的工作。

## 下周工作：

继续探索数据模型。有必要从大数据量中寻找感兴趣数据的问题入手。

查看我们的企业经济数据在经济学方面最近的分析热点。

总结现有工作。